MULTIMEDIA UNIVERSITY

STUDENT ID NO

MULTIMEDIA UNIVERSITY

FINAL EXAMINATION

TRIMESTER 3, 2017/2018

DPS5018 – INTRODUCTION TO PROBABILITY & STATISTICS

(For DIT students only)

06 JUNE 2018
09.00 am - 11.00 am
(2 Hours)

INSTRUCTIONS TO STUDENT

1. This question paper consists of **8** pages (**4** pages with **5** questions and **4** pages of appendix). Key formulae are given in the Appendix.
2. Answer **ALL** questions.
3. Write your answers in the answer booklet provided.
4. All necessary working steps must be shown.

Question 1

a) Table below shows the size distribution of families in Bangi, Selangor and the frequencies.

Size of Families	Frequency
2	34,454
3	17,525
4	15,075
5	6,863
6	2,307
7+	1,179

i) Find the total number of the families in Bangi, Selangor. (1 mark)

ii) Find the probability of the family that consist of more than three persons in their house. (2 marks)

iii) Find the percentage of the family that consist of five people and below and interpret your result. (3 marks)

b) The Women Physicians Congress compiles data on medical school faculty members of AAMC University. The contingency table below recorded the gender and rank of the faculty members.

		Gender		
		Male G_1	Female G_2	**Total**
Rank	Professor R_1	21,224	3,194	24,418
	Senior As. professor R_2	6,332	5,400	21,732
	Associate Professor R_3	25,888	14,491	40,379
	Instructor R_4	5,775	5,185	10,960
	Others R_5	781	723	1,504
	Total	70,000	28,993	98,993

i) Determine the probabilities of the events for $P(G_1)$, $P(R_1)$ and $P(R_5 \cap G_2)$. (3 marks)

ii) Find the probability of a selected member is a male given that he is an associate professor. (4 marks)

iii) Are the events R_5 and G_2 mutually exclusive? Explain your answer. (2 marks)

[TOTAL 15 MARKS]

Continued...

Question 2

a) Let X represents the number of electrical appliances sold per day at Makro Supermarket.

X	3	5	6	7	9
$P(X=x)$	0.1	0.1667	R	0.2333	0.3

i) Find the value of R. (2 marks)

ii) If the average is 6.6666, calculate the variance. (3 marks)

iii) Compute $P(3 < X \leq 9)$. (3 marks)

b) In a hospital, 40% of patients are dying of an infectious disease. If a sample of 12 patients is chosen at random, find the probability of

i) at most 7 patients are dying. (2 marks)

ii) exactly 10 patients are surviving from the disease. (3 marks)

c) The average number of calls received by a hotel reservation centre is 4 calls per minute. Find the probability of

i) less than 3 calls received in 1 minute. (2 marks)

ii) at least 5 calls received in 2 minutes. (4 marks)

d) Ministry of Women, Family and Community Development is doing a survey on the birth weight of babies in government hospital around Malaysia. Suppose we know that the birth weight of babies is normally distributed with a mean of 3500g and a standard deviation of 500g. What is the percentage that a randomly chosen new born baby that weighs between 2700g and 3800g? (6 marks)

[TOTAL 25 MARKS]

Question 3

a) M&N production manager has randomly selected a sample of 25 packs of chocolate candy and found that the mean and standard deviation of the weight of chocolate candy produced by a machine are 20.2grams and 2.3grams respectively. Construct a 90% confidence interval for the population mean. (4 marks)

b) In a random sample of 700 automobile owners, it was found that 490 prefer economic cars.

i) Find the point estimate for proportion of the owner that prefer economic cars. (1 mark)

ii) Construct a 92% confidence interval for the population proportion of those who prefer economic cars. (5 marks)

[TOTAL 10 MARKS]

Continued...

Question 4

a) A principal at a certain school claims that the students in his school are above average intelligence. A random sample of 40 students' IQ scores is recorded and has a mean score of 112. The population mean of IQ score is 100 with a standard deviation of 15.

i) State the null and alternative hypotheses. (2 marks)

ii) By using the hypothesis testing of one population mean, calculate the test statistic. (2 marks)

iii) Is there any evidence to reject the claim? Test at 5% level of significance. (4 marks)

b) Nowadays, the rate of fatal road accidents in Malaysia become crucial. The research found that the higher causes that contribute to this problem is because of the lack of people awareness for the use of seat belt. In a sample of 475 men, 198 of them are using seat belts while in a sample of 532 women, only 179 of them are using seat belts. Test the claim that men are more safety conscious than women, at 10% level of significance.

i) Determine the sample proportion of men and women who are using seat belts. (2 marks)

ii) State the null and alternative hypotheses. (2 marks)

iii) By using the hypothesis testing of two population proportions, calculate the value of the test statistic. (5 marks)

iv) Test the claim using the given level of significance and interpret your result. (4 marks)

c) Based on a research study, students need to sleep about 8 hours before attending an exam. With that, the students' mind will respond quicker and better when it is well rested. A sample of 25 students is randomly selected and found that the sample mean and standard deviation are 7.5 hours and 0.06 hours respectively. Test at 10% level of significance. (9 marks)

[TOTAL 30 MARKS]

Continued...

Question 5

a) The data in the table below shows the relationship between the number of revision hours spent per week by 8 students and the GPA that they managed to achieve in previous semester.

Hours per week, x	GPA, y
3	2.4
4	2.9
10	3.4
8	3.2
h	2.8
16	3.8
6	3.0
12	3.6

i) Given that the total revision hours spent by the students is 63 hours. Find the value of h. (2 marks)

ii) Given $\sum y = 25.1, \sum x^2 = 641$ and $\sum y^2 = 80.21$, calculate SS_{xx} and SS_{yy}. **(Give your answer in 4 decimal places)** (4 marks)

iii) Given $SS_{xy} = 13.94$, compute the coefficient of correlation, r and interpret the answer. **(Give your answer in 4 decimal places)** (3 marks)

iv) Compute b_1 and b_0 to determine the estimated least square regression. (5 marks)

v) Based on 5(a)(iv), find the revision hours spent if the student managed to score GPA 4.00 in previous semester. (2 marks)

b) Refer to part 5(a), the Dean would like to test whether the r is positively linearly correlated at 5% level of significance.

i) State the null and alternative hypotheses. (2 marks)

ii) Compute the test statistic. **(Give your answer in 4 decimal places)** (2 marks)

[TOTAL 20 MARKS]

End of Page.

APPENDIX – KEY FORMULA

- **Mean for ungrouped data**

$$\bar{x} = \frac{\sum x_i}{n} \quad \text{or} \quad \mu = \frac{\sum x_i}{N}$$

- **Mean for grouped data**

$$\bar{x} = \frac{\sum x_i f_i}{n} \quad \text{or} \quad \mu = \frac{\sum x_i f_i}{N}$$

Where x is the midpoint and f is the frequency of the class

- **Variance for ungrouped data**

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} \quad \text{or} \quad s^2 = \frac{\sum x_i^{\,2} - \frac{\left(\sum x_i\right)^2}{n}}{n-1}$$

- **Variance for grouped data**

$$s^2 = \frac{\sum (x_i - \bar{x})^2 f_i}{n-1} \quad \text{or} \quad s^2 = \frac{\sum x_i^{\,2} f_i - \frac{\left(\sum x_i f_i\right)^2}{n}}{n-1}$$

Where x is the midpoint and f is the frequency of the class

DISCRETE RANDOM VARIABLE, BINOMIAL DISTRIBUTION, AND POISSON DISTRIBUTION

- Mean of a discrete random variable x: $\mu = \sum xP(X = x)$
- Standard deviation of a discrete random variable x: $\sigma = \sqrt{\sum x^2 P(X = x) - \mu^2}$
- Binomial probability formula: $P(X = x) = \binom{n}{x} p^x q^{n-x}$
- Poisson probability formula: $P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$

- Key formula for Binomial Probability Distribution:
  - If $p \leq 0.5$
    i. $P(X = r) = B(r) - B(r-1)$
    ii. $P(X \geq r) = 1 - B(r-1)$
    iii. $P(X \leq r) = B(r)$
    iv. $P(X > r) = 1 - B(r)$
    v. $P(X < r) = B(r-1)$
    vi. $P(a \leq X \leq b) = B(b) - B(a-1)$

  - If $p > 0.5$
    i. $P(X = r) = B(n-r) - B(n-r-1)$
    ii. $P(X \geq r) = B(n-r)$
    iii. $P(X \leq r) = 1 - B(n-r-1)$
    iv. $P(a \leq X \leq b) = B(n-a) - B(n-b-1)$

- Key formula for Poisson Probability Distribution:
  - If $\mu = \lambda$
    i. $P(X = r) = Poi(r) - Poi(r-1)$
    ii. $P(X \geq r) = 1 - Poi(r-1)$
    iii. $P(X \leq r) = Poi(r)$
    iv. $P(X > r) = 1 - Poi(r)$
    v. $P(X < r) = Poi(r-1)$
    vi. $P(a \leq X \leq b) = Poi(b) - Poi(a-1)$

NORMAL AND STANDARD NORMAL PROBABILITY DISTRIBUTION

- z-value (observed value) for an x value: $Z = \frac{x - \mu}{\sigma}$

ESTIMATION

- The $(1-\alpha)100\%$ confidence interval for μ is

$\bar{x} \pm z_{\alpha/2}\sigma_{\bar{x}}$ where $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$

$\bar{x} \pm t_{\alpha/2}s_{\bar{x}}$ where $s_{\bar{x}} = \dfrac{s}{\sqrt{n}}$

- The $(1-\alpha)100\%$ confidence interval for p is

$\hat{p} \pm z_{\alpha/2}s_{\hat{p}}$ where $s_{\hat{p}} = \sqrt{\dfrac{\hat{p}\hat{q}}{n}}$

- **Sample size**

$$n = \left[\frac{\left(Z_{\alpha/2}\right)\sigma}{E}\right]^2 \qquad n = \frac{\left(Z_{\alpha/2}\right)^2 pq}{E^2}$$

HYPOTHESIS TESTING

- **ONE population**

Mean:

$$Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \quad \textbf{or} \quad t = \frac{\bar{x}-\mu}{s/\sqrt{n}}$$

Proportion:

$$Z = \frac{\hat{p}-p}{\sqrt{\dfrac{pq}{n}}}$$

- **TWO population**

Mean:

$$Z = \frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}$$

Proportion:

$$Z = \frac{(\hat{p}_1-\hat{p}_2)-(p_1-p_2)}{\sqrt{\hat{p}_p(1-\hat{p}_p)}\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}} \quad \text{where} \quad \hat{p}_p = \frac{x_1+x_2}{n_1+n_2}$$

REGRESSION

- For the **least square regression equation,** $\hat{y} = b_0 + b_1 x$

  where $b_1 = \dfrac{SS_{xy}}{SS_{xx}}$ and $b_0 = \dfrac{1}{n}\left[\sum y - b_1 \sum x\right]$

- **Correlation coefficient,** r

  $$r = \frac{SS_{xy}}{\sqrt{SS_{xx} SS_{yy}}}$$

  where

  $$SS_{xx} = \sum x^2 - \frac{\left(\sum x\right)^2}{n}, \qquad SS_{yy} = \sum y^2 - \frac{\left(\sum y\right)^2}{n} \text{ and } \qquad SS_{xy} = \sum xy - \frac{\left(\sum x\right)\left(\sum y\right)}{n}$$

- **Regression *t*-Test**

  $$t = \frac{b_1 - \beta_1}{s_b} \qquad \text{where } s_b = \frac{s_e}{\sqrt{SS_{xx}}} \qquad \text{and } s_e = \sqrt{\frac{SS_{yy} - b_1 SS_{xy}}{n-2}}$$

- **Correlation *t*-Test**

  $$t = \frac{r}{\sqrt{\dfrac{1-r^2}{n-2}}}$$